

On infinite dimensional linear programming approach to stochastic control [★]

Maryam Kamgarpour^{*} Tyler Summers^{**}

^{*} *Automatic Control Laboratory, ETH Zürich, Switzerland
(maryamk@ethz.ch)*

^{**} *Mechanical Engineering, University of Texas at Dallas
(tyler.summers@utdallas.edu)*

Abstract: We consider the infinite dimensional linear programming (inf-LP) approach for solving stochastic control problems. The inf-LP corresponding to problems with uncountable state and input spaces is in general computationally intractable. By focusing on linear systems with quadratic cost (LQG), we establish a connection between this approach and the well-known Riccati LMIs. In particular, we show that the semidefinite programs known for the LQG problem can be derived from the pair of primal and dual inf-LPs. Furthermore, we establish a connection between multi-objective and chance constraint criteria and the inf-LP formulation.

Keywords: stochastic control, linear programming, semidefinite programming

1. INTRODUCTION

Optimal control of discrete time stochastic systems can be addressed via the dynamic programming (DP) (Bellman, 1957) principle of optimality. For an infinite horizon average or discounted cost problem, the optimal cost function and control policy can be computed as the fixed point of the so-called dynamic programming operator. In general, computing this fixed point is challenging and thus, several approximate approaches based on the DP principle of optimality have been developed.

An alternative approach to solving stochastic control problems is linear programming (LP) (Puterman, 2009; Hernández-Lerma and Lasserre, 1996). If the control and input spaces are uncountable, the corresponding LP is infinite dimensional (inf-LP). In the primal form of this LP, the optimization variable is the *occupation measure*, which measures infinite horizon occupancy of state and inputs in each Borel subset of the product state input space. An optimal policy may be derived from the optimal occupation measure, while the optimal value function is the optimizer of the dual of this LP.

In addition to providing an elegant alternative formulation of the optimality conditions for a stochastic control solution, in the LP approach constraints have a natural interpretation. By properly constraining the occupation measure, one can ensure probabilistic constraints on the state trajectory or can ensure bounds on multiple objectives. Such formulations of constrained stochastic control were considered in (Borkar, 1994; Feinberg and Shwartz, 1996; Altman, 1999; Hernández-Lerma and González-Hernández, 2000; Hernández-Lerma et al., 2003).

The inf-LP formulation is in general computationally intractable. For problems with polynomial data, this inf-LP can be approximated via a sequence of semidefinite programs (SDPs) (Savorgnan et al., 2009; Summers et al., 2013). These recent works are among the few that explore the inf-LP approach for computation of optimal value function and policies in a stochastic control problem.

The abstract inf-LP work has not attempted to establish clear connections with the well known, computationally tractable Linear Matrix Inequality (LMI) formulations of optimal control. In particular, for a stochastic linear system with quadratic cost (LQG), one can formulate the so-called Riccati LMI to find the optimal value function of the LQG problem (Boyd et al., 1994; Balakrishnan and Vandenberghe, 2003). Similarly, the well known LMI formulations have not attempted to show how these results can be derived from a more general approach to stochastic optimal control, namely the inf-LP approach.

In this work, we establish the connection between the inf-LP approach and the well-known Riccati LMIs for LQG problems. This inf-LP in general, includes infinitely many constraints on the occupation measure. The relaxation of these constraints to moments up to order two of the occupation measure and taking the dual of this problem results in the well-known Riccati LMI solution approaches. Since the variables in the relaxation of primal inf-LP are discounted moments of the state and input, moment constraints or certain class of chance constraints can be naturally encoded in the inf-LP formulation.

Our paper is organized as follows. In Section 2 we review the inf-LP approach to discrete-time infinite horizon discounted stochastic control. In Section 3 we apply the approach to LQG problems. In Section 4 we provide numerical case studies. In Section 5 we summarize the results.

[★] This research is partially supported by M. Kamgarpour's European Union ERC Starting Grant, CONENE and by T. Summers' the US National Science Foundation under grant CNS-1566127.

2. INF-LP APPROACH TO STOCHASTIC CONTROL

Consider the discrete-time stochastic system

$$x_{t+1} \sim \tau(B_x | x_t, u_t), \quad (1)$$

where $x_t \in X$, $u_t \in U$, and $\tau(\cdot | x, u)$ is a stochastic kernel. It assigns a probability distribution to $B_x \in \mathcal{B}(X)$ given x and u , where $\mathcal{B}(X)$ is the set of Borel subsets of X . The stochastic control problem is defined by

$$\min_{\pi \in \Pi} \mathbb{E}_{\nu_0}^{\pi} \sum_{t=0}^{\infty} \alpha^t c_0(x_t, u_t). \quad (2)$$

Above, $c_0 : X \times U \rightarrow \mathbb{R}_+$ is the running cost and $\alpha \in (0, 1)$ is a discount factor, ν_0 is an initial state distribution. We consider randomized policies $\pi \in \Pi$, where Π is the set of probability measures on U given X . That is, for each $x \in X$, $\pi(x)$ gives a probability distribution on the input space U . The expectation \mathbb{E} is with respect to the probability measure induced by ν_0 , π and τ .

The solution to the stochastic control problem above can be characterized as the solution of an infinite dimensional linear program (inf-LP). To present this inf-LP, we first define the infinite dimensional optimization spaces for the primal and dual LPs. Define the weight functions

$$w(x, u) = \epsilon + c_0(x, u), \quad \tilde{w}(x) = \min_{u \in U} w(x, u), \quad (3)$$

where $\epsilon > 0$ so that the weights are bounded away from zero. Let $\mathcal{F}(X \times U)$, $\mathcal{F}(X)$ denote the space of real valued measurable functions with bounded w, \tilde{w} -norms, respectively. That is, for $f \in \mathcal{F}(X \times U)$, $\tilde{f} \in \mathcal{F}(X)$:

$$\sup_{(x,u)} \frac{|f(x,u)|}{w(x,u)} < \infty, \quad \sup_x \frac{|\tilde{f}(x)|}{\tilde{w}(x)} < \infty.$$

Let $\mathcal{M}(X \times U)$, $\mathcal{M}(X)$ denote the space of measures with finite w, \tilde{w} -variations, respectively. That is, for $\mu \in \mathcal{M}(X \times U)$, $\tilde{\mu} \in \mathcal{M}(X)$:

$$\int_{X \times U} w d\mu < \infty, \quad \int_X \tilde{w} d\tilde{\mu} < \infty. \quad (4)$$

Define the linear map $T : \mathcal{M}(X \times U) \rightarrow \mathcal{M}(X)$ as:

$$[T\mu](B) = \tilde{\mu}(B) - \alpha \int_{X \times U} \tau(B | x, u) \mu(dx, du), \quad (5)$$

where $\tilde{\mu}(B) := \mu(B, U)$ and $B \in \mathcal{B}(X)$. Analogously, define the linear map $T^* : \mathcal{F}(X) \rightarrow \mathcal{F}(X \times U)$ as:

$$[T^*v](x, u) = v(x) - \alpha \int_X \tau(dy | x, u) v(y).$$

Note that the second term above $\int_X \tau(dy | x, u) v(y)$, is the expectation of the function v under the stochastic kernel τ . One can verify that T and T^* are adjoint operators:

$$\langle T^*v, \mu \rangle_{X \times U} = \langle v, T\mu \rangle_X,$$

where the bilinear maps are given by:

$$\begin{aligned} \langle c, \mu \rangle_{X \times U} &= \int_{X \times U} c(x, u) \mu(dx, du), \\ \langle v, \nu \rangle_X &= \int_X v(x) \nu(dx). \end{aligned}$$

In the remainder, for simplicity, we drop the subscript of $\langle \cdot, \cdot \rangle$ since the space is clear from the context. To formulate the inf-LP corresponding to stochastic control, we need the following standard assumptions (Hernández-Lerma and Lasserre, 1996).

Assumption 1.

- (a) The cost c_0 is lower semi-continuous and inf-compact, that is, for every $x \in X$, $r \in \mathbb{R}$, the set $\{u \in U \mid c_0(x, u) \leq r\}$ is non-empty and compact.
- (b) The stochastic kernel τ is weakly continuous.
- (c) $\sup_{X \times U} \int_X \tilde{w}(y) \tau(dy | x, u) / w(x, u) < \infty$.
- (d) $\nu_0 \in \mathcal{M}_+(X)$.

Let $\mathcal{M}_+(X \times U) \subset \mathcal{M}(X \times U)$ denote the cone of non-negative measures. For $\nu_0 \in \mathcal{M}_+(X)$, the constraint on $\mu \in \mathcal{M}(X \times U)$, denoted by $\nu_0 - T\mu = 0$ refers to

$$\nu_0(B_x) - [T\mu](B_x) = 0, \quad \forall B_x \in \mathcal{B}(X). \quad (6)$$

Theorem 1. The stochastic control problem (1), (2) can be equivalently formulated as the following inf-LP:

$$\min_{\mu \in \mathcal{M}_+(X \times U)} \langle c_0, \mu \rangle \quad (\text{P-SC})$$

$$\text{s.t.} \quad \nu_0 - T\mu = 0. \quad (7)$$

We summarize the idea of the proof and refer the readers to (Hernández-Lerma and Lasserre, 1996) for details. Given a policy $\pi \in \Pi$, one can define $\mu \in \mathcal{M}_+(X \times U)$ as

$$\mu(B_x, B_u) = \sum_{t=0}^{\infty} \alpha^t \mathbb{P}_{\nu_0}^{\pi} \{(x_t, u_t) \in (B_x, B_u)\}, \quad (8)$$

where $B_x \in \mathcal{B}(X)$, $B_u \in \mathcal{B}(U)$. This measure corresponds to discounted probability of (x_t, u_t) being in any Borel subset of $X \times U$ and is referred to as the occupation measure. It can be verified that the occupation measure satisfies $\nu_0 - T\mu = 0$. Furthermore, given any $\mu \in \mathcal{M}_+(X \times U)$, there exists a policy $\varphi \in \Pi$, satisfying

$$\mu(B_x, B_u) = \int_{B_x} \varphi(B_u | x) \tilde{\mu}(dx), \quad (9)$$

for all $B_x \in \mathcal{B}(X)$, $B_u \in \mathcal{B}(U)$ [Proposition D.8(a) in (Hernández-Lerma and Lasserre, 1996)]. It can be shown that the cost (2) corresponding to the policy φ is

$$\mathbb{E}_{\nu_0}^{\varphi} \sum_{t=0}^{\infty} \alpha^t c_0(x_t, u_t) = \langle c_0, \mu \rangle. \quad (10)$$

Putting the above results together, the problem of finding the optimal policy for (2) can be equivalently formulated as finding a measure minimizing (10) subject to (7).

Whereas the inf-LP above provides the optimal occupation measure and the optimal policy for the stochastic control problem, the dual of this inf-LP can be used to find the optimal value function. Furthermore, the duality gap is zero (Hernández-Lerma and Lasserre, 1996).

To define this dual inf-LP, let the constraint on $v \in \mathcal{F}(X)$, denoted by $c_0 - T^*v \geq 0$ refer to

$$c_0(x, u) - [T^*v](x, u) \geq 0, \quad \forall (x, u) \in X \times U. \quad (11)$$

The dual inf-LP is given by:

$$\max_{v \in \mathcal{F}(X)} \langle v, \nu_0 \rangle \quad (\text{D-SC})$$

$$\text{s.t.} \quad c_0 - T^*v \geq 0. \quad (12)$$

Remark. Constraint (12) is the Bellman inequality. In particular, based on the Bellman principle of optimality, a function v^* is the optimal value function of the stochastic control if and only if $c_0 - T^*v = 0$. Thus, the optimizer of the above inf-LP satisfies the Bellman equality.

3. INF-LP APPROACH TO LQG PROBLEMS

Consider the linear system as a specialization of (1):

$$x_{t+1} = Ax_t + Bu_t + \omega_t, \quad (13)$$

where $x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$, $\omega \in \mathbb{R}^n$, ω_t , are independent identically distributed Gaussian random variables and for all t , $E\{\omega_t\} = 0$ and $E\{\omega_t \omega_t^T\} = W$. The initial state is independent of the stochastic noise and has a distribution ν_0 , with mean $E\{x_0\} = m_0$ and covariance $E\{x_0 x_0^T\} = \Sigma_0$. The discounted linear quadratic Gaussian (LQG) problem is formulated as:

$$\min_{\pi \in \Pi} \mathbb{E}_{\nu_0}^{\pi} \sum_{t=0}^{\infty} \alpha^t (x_t^T Q_0 x_t + u_t^T R_0 u_t). \quad (14)$$

We assume the pair (A, B) is controllable and the pair (A, C) is observable, where $Q_0 = C^T C$. Denote by \mathcal{S}^n , \mathcal{S}_{++}^n and \mathcal{S}_{++}^n the set of $n \times n$ symmetric, symmetric positive semidefinite and symmetric positive definite matrices, respectively. We assume $Q_0 \in \mathcal{S}_{++}^n$ and $R_0 \in \mathcal{S}_{++}^m$.

To apply the inf-LP approach to this problem, first we verify Assumption (1) as follows. The weight functions (3) in the LQG problem are $w(x, u) = \epsilon + x^T Q_0 x + u^T R_0 u$, $\tilde{w}(x) = \epsilon + x^T Q_0 x$. Thus, $\mathcal{F}(X \times U)$ and $\mathcal{F}(X)$ are spaces of functions over $X \times U = \mathbb{R}^{n \times m}$, $X = \mathbb{R}^n$ respectively, that do not grow faster than quadratic functions. Furthermore, by definition (4), $\mathcal{M}(\mathbb{R}^{n \times m})$, $\mathcal{M}(\mathbb{R}^n)$ are sets of measures that have bounded variance. From $Q_0 \in \mathcal{S}_{++}^n$, $R_0 \in \mathcal{S}_{++}^m$, part (a) of Assumptions (1) holds. The stochastic kernel τ is Gaussian, which is continuous and has finite variance, satisfying part (b). Part (c) holds since

$$\begin{aligned} \int_X w_0(y) \tau(dy|x, u) &= \epsilon + x^T A^T Q_0 A x \\ &+ 2x^T A^T Q_0 B u + u^T B^T Q_0 B u + \text{Tr}(Q_0 W) \in \mathcal{F}(\mathbb{R}^{n \times m}). \end{aligned}$$

Finally, part (d) holds due to finite variance of the initial state distribution ν_0 .

3.1 Primal and Dual SDPs

We consider a relaxation of the inf-LP (P-SC), resulting in an equivalent restriction of (D-SC), to obtain a tractable formulation of these inf-LPs. These formulations are then connected to the well-known Riccati LMIs to solve the LQG (Balakrishnan and Vandenberghe, 2003).

First, consider the constraint $\nu_0 - T\mu = 0$ in (P-SC). It can be verified that this is equivalent to $\langle v, \nu_0 - T\mu \rangle = 0$, $\forall v \in \mathcal{F}(X)$. We relax this constraint by restricting \mathcal{F} to a subset $\hat{\mathcal{F}}(X)$. In particular, define

$$\hat{\mathcal{F}}(X) = \{v \in \mathcal{F}(X) \mid v(x) = x^T P x + q^T x + r\}, \quad (15)$$

$P \in \mathcal{S}^n$, $q \in \mathbb{R}^n$, $r \in \mathbb{R}$. Since $v \in \mathcal{F}(X)$ are quadratic, the infinitely many constraint (7) on the measure μ are relaxed to a set of finite constraints on its moments of order up to two. These constraints will be derived as follows.

Introduce moments of measure μ as:

$$\begin{aligned} m &= \int_{X \times U} \mu(dx, du) = \mu(X \times U) \in \mathbb{R}_+, \\ m_x &= \int_X x \mu(dx, U) \in \mathbb{R}^n, \\ Z_{xx} &= \int_X x x^T \mu(dx, U) \in \mathcal{S}_{++}^n. \end{aligned}$$

Similarly, m_u, Z_{xu}, Z_{uu} are defined. For any $S \in \mathbb{R}^{n \times n}$, $\text{Tr}(S)$ denotes trace of the matrix S .

Proposition 2. (Primal SDP for LQG). By constraining $\mathcal{F}(X)$ to $\hat{\mathcal{F}}(X)$, we obtain a relaxation of (P-SC) as

$$\begin{aligned} \min \quad & \text{Tr}(Q_0 Z_{xx}) + \text{Tr}(R_0 Z_{uu}) \quad (\text{P-LQG}) \\ \text{s.t.} \quad & \Sigma_0 + m_0 m_0^T - Z_{xx} + \alpha A Z_{xx} A^T + A Z_{xu} B^T \\ & + B Z_{xu}^T A^T + B Z_{uu} B^T + m W = 0_{n \times n} \quad (\text{C2}) \\ & m_0 - m_x + \alpha(A m_x + B m_u) = 0_{n \times 1}, \quad (\text{C1}) \\ & \alpha m - m + 1 = 0, \quad (\text{C0}) \\ & Z := \begin{bmatrix} m & m_x^T & m_u^T \\ m_x & Z_{xx} & Z_{xu} \\ m_u & Z_{xu}^T & Z_{uu} \end{bmatrix} \succeq 0, \quad (\text{Cp}) \end{aligned}$$

over the variables $(m, m_x, m_u, Z_{xx}, Z_{xu}, Z_{uu})$.

Proof. Based on the definition of the moments, the cost functions in Problem (14) can be expressed as:

$$\langle c_0, \mu \rangle = \text{Tr}(Q_0 Z_{xx}) + \text{Tr}(R_0 Z_{uu}).$$

Next, expanding $\langle v, \nu_0 - T\mu \rangle = 0$ we obtain the first term as

$$\begin{aligned} \langle v, \nu_0 \rangle &= \langle x^T P x + q^T x + r, \nu_0 \rangle \\ &= \text{Tr}(P \Sigma_0) + m_0^T P m_0 + q^T m_0 + r. \end{aligned}$$

Using definition of T in (5), we expand $\langle v, -T\mu \rangle$. The first term is:

$$\begin{aligned} \langle v, \tilde{\mu} \rangle &= \langle x^T P x + q^T x + r, -\tilde{\mu} \rangle = \\ &= - \int_X v(x) \mu(dx, U) = -\text{Tr}(P Z_{xx}) - q^T m_x - m r. \end{aligned}$$

The second term is obtained as:

$$\begin{aligned} \alpha \times (\text{Tr}(P(AZ_{xx}A^T + AZ_{xu}B^T + BZ_{xu}^TA^T + BZ_{uu}B^T)) \\ + q^T(Am_x + Bm_u) + m(\text{Tr}(PW) + r)). \end{aligned}$$

In the above, we used the fact that the measure $\tau(dx|y, u)$ has mean $Ay + Bu$ and covariance W . Each of the terms in the constraint expansion above are linear in the variables P, q, r . Since $\langle v, \nu_0 - T\mu \rangle = 0$ must hold for all $v \in \hat{\mathcal{F}}(X)$, that is, for all $P \in \mathcal{S}^n, q \in \mathbb{R}^n, r \in \mathbb{R}$, the corresponding coefficients of these variables need to equal zero. From this, we obtain the set of affine constraints, (C0), (C1), (C2). Constraint (Cp) holds since Z is moment of a positive measure μ (Lasserre, 2009). \square

Similarly, we can obtain the dual SDP as follows.

Proposition 3. (Dual SDP for LQG). By constraining $\hat{\mathcal{F}}(X)$ to $\mathcal{F}(X)$, we obtain a restriction of (D-SC) as follows:

$$\begin{aligned} \min \quad & \text{Tr}(P \Sigma_0) + \text{Tr}(P m_0 m_0^T) + q^T m_0 + r \quad (\text{D-LQG}) \\ \text{s.t.} \quad & \begin{bmatrix} s_0 & s_1^T & s_2^T \\ s_1 & S_{11} & S_{12} \\ s_2 & S_{12}^T & S_{22} \end{bmatrix} \succeq 0, \end{aligned}$$

with optimization variables, P, q, r and

$$\begin{aligned} s_0 &= r(\alpha - 1) + \alpha \text{Tr}(PW), \\ s_1 &= \frac{1}{2}(-I + \alpha A^T)q, \quad s_2 = \frac{\alpha}{2}B^T q, \\ \begin{bmatrix} S_{11} & S_{12} \\ S_{12}^T & S_{22} \end{bmatrix} &= \begin{bmatrix} \alpha A^T P A - P + Q_0 & \alpha A^T P B \\ \alpha B^T P A & R_0 + \alpha B^T P B \end{bmatrix}. \end{aligned}$$

Furthermore, this SDP is the dual of (P-LQG).

Proof. The term $\langle v, \nu_0 \rangle$ in the cost function was discussed in Proof of Proposition (2). For $v \in \hat{\mathcal{F}}(X)$, Constraint (11) becomes

$$\begin{aligned} & x^T(-P + \alpha A^T P A + Q_0)x + 2\alpha x^T A^T P B u \\ & + u^T(\alpha B^T P B + R_0)u + q^T(-I + \alpha A)x \\ & + \alpha q^T B u + r(\alpha - 1) + \alpha \text{Tr}(P W) \geq 0, \quad \forall (x, u) \in X \times U. \end{aligned}$$

An equivalent way of writing the above constraint is:

$$\begin{bmatrix} 1 \\ x \\ a \end{bmatrix}^T \begin{bmatrix} s_0 & s_1 & s_2 \\ s_1^T & S_{11} & S_{12} \\ s_2^T & S_{12}^T & S_{22} \end{bmatrix} \begin{bmatrix} 1 \\ x \\ a \end{bmatrix} \succeq 0,$$

which leads to the constraint in (D-LQG). Using SDP duality (Vandenberghe and Boyd, 1996), it can also be verified that (D-LQG) is dual of (P-LQG). \square

Remark. The primal and dual SDPs above are a generalization of the existing results in literature due to the additional terms arising from zero and first order moments of μ . If $m_0 = 0$ from controllability of the pair (A, B) we have $m_x = 0$, $m_u = 0$. Thus, removing the corresponding dual variable $q \in \mathbb{R}^n$, we can obtain that $r = \frac{\alpha}{1-\alpha} \text{Tr}(P W)$. This leads to the standard results in (Willems, 1971; Boyd et al., 1994; Balakrishnan and Vandenberghe, 2003):

$$\min \text{Tr}(P \Sigma_0) + \frac{\alpha}{1-\alpha} \text{Tr}(P W) \quad (\text{D-LQG0})$$

$$\text{s.t.} \quad S := \begin{bmatrix} A^T P A - P + Q_0 & \alpha A^T P B \\ \alpha B^T P A & R_0 + \alpha B^T P B \end{bmatrix} \succeq 0.$$

In the rest of the paper, we consider $m_0 = 0$ and thus, we work with (D-LQG0) and its dual.

If the SDP (D-LQG0) and its dual have non-empty optimal sets, the complementary slackness condition holds (Vandenberghe and Boyd, 1996):

$$\begin{aligned} Z^* S^* = 0 & \iff \begin{bmatrix} Z_{xx} & Z_{xu} \\ Z_{xu}^T & Z_{uu} \end{bmatrix} \times \\ & \begin{bmatrix} -P + Q_0 + \alpha A^T P A & \alpha A^T P B \\ \alpha B^T P A & R_0 + \alpha B^T P B \end{bmatrix} = 0, \end{aligned}$$

where we dropped $*$ from individual terms above. Expanding above equality, we obtain that

$$0 = -P + Q + \alpha A^T P A + \alpha A^T P B (R + \alpha B^T P B)^{-1} \alpha B^T P A, \quad (16)$$

$$Z_{xu}^T Z_{xx}^{-1} = (R + \alpha B^T P B)^{-1} \alpha B^T P A, \quad (17)$$

$$Z_{uu} = Z_{xu}^T Z_{xx}^{-1} Z_{xu}. \quad (18)$$

Equation (16) is the algebraic Riccati equation of the infinite horizon discounted LQG problem. Equation (17) provides the optimal controller gain, $K = Z_{xu}^T Z_{xx}^{-1}$.

Remark. An alternative derivation of the optimal policy is provided by considering the occupation measure μ in the inf-LP. Since $\frac{1}{m} \mu(X \times U) = (1-\alpha) \mu(X \times U)$ is a Gaussian measure (alternatively, by considering only the knowledge of the first and second order moments of this measure), the conditional measure $\varphi(u|x)$ (9) can be obtained as

$$\varphi(u|x) \sim \mathcal{N}(m_{u|x}, Z_{u|x}),$$

with the conditional mean $m_{u|x} = m_u + Z_{xu}^T Z_{xx}^{-1} x$ and covariance $Z_{u|x} = \frac{1}{m} (Z_{uu} - Z_{xu}^T Z_{xx}^{-1} Z_{xu})$. By complementary slackness of (18), the covariance of this measure is zero and thus, the optimal policy predicted by inf-LP (P-SC) is deterministic and is equal to $\varphi(x) = Z_{xu}^T Z_{xx}^{-1} x$.

3.2 Constrained LQG

One of the advantages of the inf-LP (P-SC) is that constraints of the form $\mathbb{E}_{\nu_0}^{\pi} \sum_{t=0}^{\infty} \alpha^t c_i(x_t, u_t) \leq \beta_i$, $i = 1, 2, \dots, N$, can readily be incorporated. In particular, from the definition of the occupation measure (8):

$$\mathbb{E}_{\nu_0}^{\pi} \sum_{t=0}^{\infty} \alpha^t c_i(x_t, u_t) \leq \beta_i \iff \langle c_i, \mu \rangle \leq \beta_i.$$

Such constraints correspond to multi-objective stochastic control, where c_i , for $i = 1, \dots, N$, denotes a set of additional objectives and $\beta_i \in \mathbb{R}_+$ are desired bounds.

For the LQG problem, let $c_1(x_t, u_t) = x_t^T Q_1 x_t + u_t^T R_1 u_t$. Then, constraint $\langle c_1, \mu \rangle \leq \beta_1$ is equivalent to $\text{Tr}(Q_1 Z_{xx}) + \text{Tr}(R_1 Z_{uu}) \leq \beta_1$ in the primal SDP (P-LQG). From our derivation of (D-LQG0), it can be verified that the dual SDP with the additional constraint is

$$\min \text{Tr}(P \Sigma_0) + \frac{\alpha}{1-\alpha} \text{Tr}(P W) - \gamma \beta \quad (\text{C-LQG})$$

$$\text{s.t.} \quad \begin{bmatrix} -P + Q + \alpha A^T P A & \alpha A^T P B \\ \alpha B^T P A & R + \alpha B^T P B \end{bmatrix} \succeq 0,$$

where $Q = Q_0 + \beta Q_1$ and $R = R_0 + \beta R_1$ and the optimization variables are P and the dual multiplier of the constraint $\gamma > 0$. This is consistent with alternative derivations in multi-criterion LQG in (Boyd et al., 1994).

Due to Z_{xx} and Z_{uu} corresponding to the second order discounted moments of the occupation measure, constraints $\text{Tr}(Q_1 Z_{xx}) \leq \beta_1$ can also be used to pose chance constraints on the state of the form

$$\sum_{t=0}^{\infty} \alpha^t \mathbb{P}_{\nu_0}^{\pi}(|g^T x_t| < h) \geq 1 - \epsilon. \quad (19)$$

Proposition 4. Given a policy π , let μ be the resulting occupation measure. Let $Q_1 = gg^T$, and $\beta_1 = \epsilon h^2$

$$\text{Tr}(Q_1 Z_{xx}) \leq \beta_1 \Rightarrow \sum_{t=0}^{\infty} \alpha^t \mathbb{P}_{\nu_0}^{\pi}(|g^T x_t| \leq h) \geq 1 - \epsilon.$$

Furthermore, for the case in which $\pi(x) = Kx$ and in the LQG setting, letting $\beta_1 = \frac{h^2}{2(\text{erf}^{-1}(1-\epsilon/m))^2}$, where erf denotes the error function, the constraint above is also sufficient for the chance constraint.

Proof. Define $X_h = \{x : |g^T x| \geq h\}$, so that $\mu(X_h, U) = \sum_{t=0}^{\infty} \alpha^t \mathbb{P}_{\nu_0}^{\pi}(x_t \in X_h)$. Then, the chance constraint (19) can be written as $\mu(X_h, U) \leq \epsilon$. Now

$$0 \leq h^2 \mathbf{1}_{X_h} \leq (g^T x)^2 \mathbf{1}_{X_h}$$

Taking integral with respect to μ we obtain that

$$h^2 \mu(X_h, U) \leq \int_X (g^T x)^2 \mathbf{1}_{X_h} \mu(dx, U) \leq g^T Z_{xx} g.$$

It follows that $\frac{g^T Z_{xx} g}{h^2} < \epsilon \Rightarrow \mu(X_h, U) < \epsilon$.

If π is linear, the resulting occupation measure μ is a scaled (by factor $m = \frac{1}{1-\alpha}$) Gaussian measure. Since the cumulative distribution function of the Gaussian distribution is invertible and is given through the erf function, we have

$$\begin{aligned} \mu(X_h, U) < \epsilon & \iff 1 + \text{erf}\left(\frac{-h}{\sigma\sqrt{2}}\right) \leq \frac{\epsilon}{m} \\ & \iff \frac{h^2}{2(\text{erf}^{-1}(1-\epsilon/m))^2} \geq g^T Z_{xx} g. \quad \square \end{aligned}$$

Remark. The above chance constraints were also considered in infinite horizon average cost LQG problems (Schildbach et al., 2015). The authors derived analogous SDPs for the average cost criterion. Their approach was through considering the steady-state second order moments of the closed loop linear system, corresponding to a linear policy. As shown here, this approach is equivalent to the relaxation of the primal infinite dimensional LP from which the occupation measure and the corresponding second order moments were derived.

4. NUMERICAL CASE STUDIES

Our goal is to use the constrained LQG formulation of the previous section to study effects of multi-objective and chance constraints on the LQG problem. To this end, we consider two linear systems, each with a nominal objective, and closed-loop infinite horizon second order moment constraints on states or inputs. In both cases, we solved SDP (P-LQG) using the parser CVX (Grant et al., 2008) with the solver SDPT3 (Toh et al., 1999).

Our first example is a second order system to study effects of chance constraints (19). In our second example, we consider a model for a miniature coaxial helicopter linearized around a hover maneuver, with a nominal objective of minimizing deviations from hover. Our secondary objective is to minimize control energy.

4.1 Two-state system with state constraints

The system dynamics parameters are

$$A = \begin{bmatrix} 1 & 0.1 \\ 0 & 1 \end{bmatrix}, B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, W = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix}.$$

The primary and secondary objective parameters are

$$Q_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, Q_1 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, R_0 = 1, R_1 = 0.$$

The discount factor is $\alpha = 0.99$ and $x_0 \sim \mathcal{N}(m_0, \Sigma_0)$, with mean $m_0 = [-0.46, 0.58]^T$ and covariance $\Sigma_0 = I$. The second objective, $\text{Tr}(Q_1 Z_{xx})$ is constrained to be less than a parameter β . As such, we require $\sum_{t=0}^{\infty} \alpha^t P_{v_0}^{\pi}([0, 1]x_t \leq h) \geq 1 - \epsilon$. This has the interpretation of a soft constraint for the second state to remain close to zero.

We vary β between the value of the second objective achieved using the optimal policy for first objective without the constraint and a lower tighter value that forces the constraint to be active. Figure (4.1) shows the optimal state covariance Z_{xx}^* . It is seen that as the constraint is tightened, the discounted occupancy ellipse changes to one with less variation in the second state, due to definition of Q_1 , but more variation in the first state. The cost is 378 without the constraint and 981 with $\beta = 15$.

4.2 Miniature coaxial helicopter

We now consider a simplified eight-state model of a miniature two-rotor coaxial helicopter linearized around a hover maneuver based on (Kunz et al., 2013; Summers et al., 2013). The states of the system are the three-dimensional position and heading deviations from a desired hover pose in an inertial reference frame and the associated velocities in a body reference frame. There are four inputs: pitch,

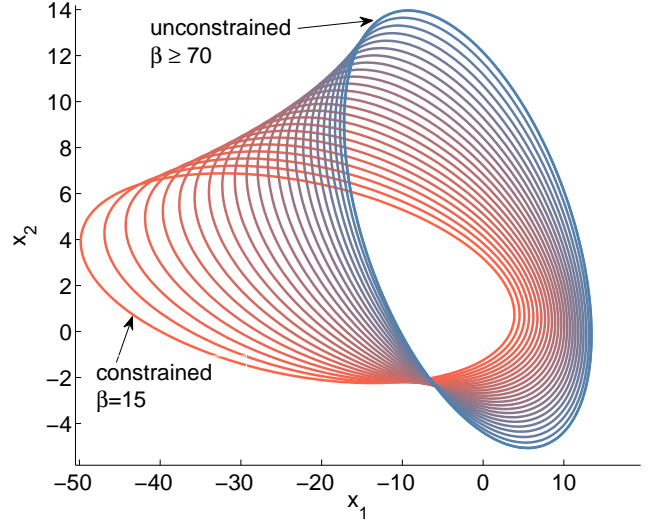


Fig. 1. Discounted state occupancy ellipses $\{x \in \mathbf{R}^2 \mid (x - m_x^*)^T Z_{xx}^{*-1} (x - m_x^*) = 1\}$ as the constraint associated with the secondary cost is tightened from unconstrained (blue) to $\beta = 15$ (red).

roll, thrust, and yaw, used for forward flight, sideways flight, vertical flight, and heading change, respectively. Pitch and roll are actuated with a swashplate mechanism connected to two servos. Thrust is actuated by the rotational speed of the rotor motors, and yaw is actuated by a rotational speed difference of the rotor motors. The pitch and roll angles and velocities are neglected in the model, and the pitch and roll inputs are assumed to act directly on the lateral position states.

The dynamics are discretized in time by Euler integration with sampling time t_s . The parameters of the the discrete-time system dynamics are

$$A = \begin{bmatrix} I_4 & t_s I_4 \\ 0 & I_4 + t_s \text{diag}([k_x, k_y, k_z, k_\psi]) \end{bmatrix}, \\ B = \begin{bmatrix} 0_4 \\ t_s \text{diag}([b_x, b_y, b_z, b_\psi]) \end{bmatrix}, W = \begin{bmatrix} 0_4 & 0 \\ 0 & 0.1 I_4 \end{bmatrix},$$

where $[k_x, k_y, k_z, k_\psi] = [-0.5, -0.5, 0, -5]$ represent fuselage drag parameters and $[b_x, b_y, b_z, b_\psi] = [2.0, 2.1, 11, 18]$ represent inertial parameters mapping actuator influence to state derivatives. The values of the parameters are taken from (Kunz et al., 2013) and are based on a grey-box system identification with experimental data.

We consider a scenario in which the deviations from the desired hover pose are to be minimized, subject to an infinite-horizon closed-loop constraint on the expected discounted control energy. The trade-off between trajectory optimization and energy cost minimization is a classical control tradeoff. It is encoded in our framework with the primary and secondary cost parameters

$$Q_0 = I_8, Q_1 = 0_8, R_0 = 0_4, R_1 = I_4.$$

The discount factor is $\alpha = 0.99$ and the threshold is $\beta = 50$. The constraint on control energy is achieved with the constrained LQG formulation by solving (P-LQG). This constraint is satisfied at a price of poorer regulation compared to the unconstrained case as illustrated by the closed-loop system performance in Fig. 2.

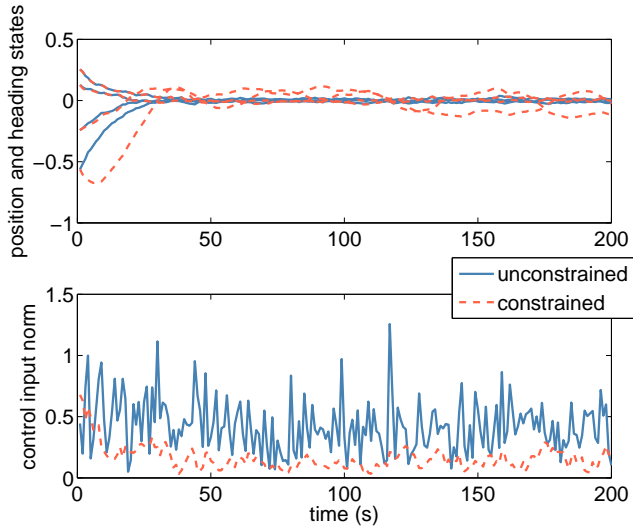


Fig. 2. Classical state regulation and control energy trade-off achieved with the constrained LQG formulation. The plots show a realization of the position and heading state and control norm evolution with and without the constraint on the closed-loop infinite-horizon expected discounted control energy.

5. CONCLUSIONS

We established a link between the semidefinite programs (SDPs) for solving the LQG problem and the infinite dimensional linear programming (inf-LP) approach to stochastic control. The inf-LP approach is an equivalent alternative to the Dynamic Programming principle of optimality. While the inf-LP formulation has been known since 1950's, its computational aspects and connections with existing control theoretic results have not been fully explored. We showed that the LMI derived from the occupation measure formulation of the inf-LP corresponds to the dual of the well-known Riccati LMI. Furthermore, given second order moments of the occupation measure, we showed that multi objective and chance constraints have a natural interpretation in this framework, and these formulations coincide with alternative approaches to derive the results. We illustrated the constrained LQG problem with two numerical case studies.

Extensions of this work to continuous-time LQG, and average cost LQG problems are straightforward and will complete the picture. While a rich theory of approximate dynamic programming (ADP) exists, it would be interesting to enrich the approximation procedures, through advanced optimization techniques for solving the inf-LPs corresponding to stochastic control. There has been recent promising steps towards this objective (??). It will be interesting to further apply these techniques to large-scale and constrained stochastic control problems.

REFERENCES

Altman, E. (1999). *Constrained Markov decision processes*, volume 7. CRC Press.

Balakrishnan, V. and Vandenberghe, L. (2003). Semidefinite programming duality and linear time-invariant systems. *Automatic Control, IEEE Transactions on*, 48(1), 30–41.

Bellman, R.E. (1957). *Dynamic Programming*. Princeton University Press.

Borkar, V.S. (1994). Ergodic control of Markov chains with constraints-the general case. *SIAM Journal on Control and Optimization*, 32(1), 176–186.

Boyd, S.P., El Ghaoui, L., Feron, E., and Balakrishnan, V. (1994). *Linear matrix inequalities in system and control theory*, volume 15. SIAM.

Feinberg, E.A. and Schwartz, A. (1996). Constrained discounted dynamic programming. *Mathematics of Operations Research*, 21(4), 922–945.

Grant, M., Boyd, S., and Ye, Y. (2008). CVX: Matlab software for disciplined convex programming.

Hernández-Lerma, O. and González-Hernández, J. (2000). Constrained Markov control processes in Borel spaces: the discounted case. *Mathematical Methods of Operations Research*, 52(2), 271–285.

Hernández-Lerma, O., González-Hernández, J., and López-Martínez, R.R. (2003). Constrained average cost Markov control processes in Borel spaces. *SIAM Journal on Control and Optimization*, 42(2), 442–468.

Hernández-Lerma, O. and Lasserre, J.B. (1996). *Discrete-time Markov control processes*. Springer.

Kunz, K., Huck, S., and Summers, T. (2013). Fast model predictive control of miniature helicopters. In *European Control Conference*, 1377–1382. IEEE.

Lasserre, J.B. (2009). *Moments, positive polynomials and their applications*, volume 1. World Scientific.

Puterman, M.L. (2009). *Markov decision processes: discrete stochastic dynamic programming*, volume 414. John Wiley & Sons.

Savorgnan, C., Lasserre, J.B., and Diehl, M. (2009). Discrete-time stochastic optimal control via occupation measures and moment relaxations. In *Proceedings of the 48th IEEE Conference on Decision and Control held jointly with the 28th Chinese Control Conference*, 519–524.

Schildbach, G., Goulart, P., and Morari, M. (2015). Linear controller design for chance constrained systems. *Automatica*, 51, 278–284.

Summers, T.H., Kunz, K., Kariotoglou, N., Kamgarpour, M., Summers, S., and Lygeros, J. (2013). Approximate dynamic programming via sum of squares programming. In *Control Conference (ECC), 2013 European*, 191–197. IEEE.

Toh, K.C., Todd, M.J., and Tütüncü, R.H. (1999). SDPT3? a MATLAB software package for semidefinite programming, version 1.3. *Optimization methods and software*, 11(1-4), 545–581.

Vandenberghe, L. and Boyd, S. (1996). Semidefinite programming. *SIAM review*, 38(1), 49–95.

Willems, J. (1971). Least squares stationary optimal control and the algebraic riccati equation. *IEEE Transactions on Automatic Control*, 16(6), 621–634.